

Citation for published version:

Petropoulos, F 2019, -Reflections from the Next Generation of Forecasters. in K Nikolopoulos & DD Thomakos (eds), *Forecasting With The Theta Method: Theory and Applications*. Wiley, pp. 163-176.

Publication date:

2019

Document Version

Early version, also known as pre-print

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

θ -Reflections from the Next Generation of Forecasters¹

Fotios Petropoulos

School of Management, University of Bath, UK

Abstract

The theta method, in its simplest implementation as applied by Assimakopoulos and Nikolopoulos (2000), had a robust performance in a wide variety of data. This chapter investigates ways to expand the theta method and further improve its forecasting performance. We argue that theta is a framework rather than a method. For example, different or more theta lines may be considered, theta lines may be extrapolated by alternative forecasting methods or unequal combinations of the forecasts may be calculated. We propose a “robust” theta method that incorporates a theta line representing the seasonally-adjusted data and extrapolated by Damped exponential smoothing as well as a short-term linear trend line. Finally, we explore the various open-source theta implementations.

1 Introduction

The theta method, as it was applied by Assimakopoulos and Nikolopoulos (2000) to produce forecasts for the M3-Competition (Makridakis and Hibon, 2000), involved several ad-hoc decisions and simplifications, such as

- adjust the data for seasonality using multiplicative classical decomposition,
- decompose the seasonally adjusted data into exactly two theta lines, with theta coefficients equal to 0 and 2 corresponding to the simple linear regression line on time and a line with double the curvatures of the seasonally adjusted data respectively,
- extrapolate the theta line 0 as usual, using the regression model fitted on that line the assuming a linear trend,

¹Pre-print version of the chapter submitted in K. Nikolopoulos & D.D. Thomakos, *Forecasting With The Theta Method: Theory and Applications*, published by Wiley.

Email address: f.petropoulos@bath.ac.uk (Fotios Petropoulos)

- extrapolate the theta line 2 using the simplest form of exponential smoothing models, Simple (or Single) Exponential Smoothing (SES), and
- combine the forecasts produced for theta lines 0 and 2 using equal weights.

We will further on refer to the theta method based on the aforementioned settings as the *standard theta method*. The standard theta method can be extended by considering several deviations from the standard set-up, such as

- alternative seasonal adjustments,
- different values for the theta parameters,
- multiple theta lines,
- alternative extrapolation methods, and
- unequal combination weights.

This chapter reports some results in the literature as well as some new results that show the potential of extending the standard theta method. In doing so, the next section describes the data used to produce the results as well as how performance is measured. The last section reviews the available open source packages and functions for the R statistical software that can be used to apply the theta method in practice.

2 Design

The results reported in this chapter have been produced using the monthly subset of the M3-Competition data set (Makridakis and Hibon, 2000). This consists of 1428 real monthly time series with varying in-sample lengths, from 48 periods (4 years) to 126 periods (10.5 years). The length of the out-of-sample period is 18 months for all time series. In other words, forecasts for the next 18 periods are produced and tested for their accuracy. The monthly series of the M3-Competition originate from various fields, such as demographic, finance, industry, macro, micro and other.

Forecasting performance is measured by means of the symmetric Mean Absolute Percentage Error (sMAPE). While this error measure has some drawbacks (Goodwin and Lawton, 1999), we opt to use this measure instead of other alternatives as to be directly comparable to the original accuracy results published by Makridakis and Hibon (2000). The sMAPE for one time series is

calculated as

$$\text{sMAPE} = \frac{200}{H} \sum_{h=1}^H \frac{|y_{n+h} - f_{n+h|n}|}{|y_{n+h}| + |f_{n+h|n}|}, \quad (1)$$

where n is the number of observations in the in-sample, y_{n+h} is the actual value at period $n+h$, $f_{n+h|n}$ is the corresponding forecast as produced at origin n and H is the maximum forecast horizon (18 months). Given that this error measure is scale independent, the values of sMAPE can be summarised across several series.

3 Seasonal adjustment

In their original implementation, Assimakopoulos and Nikolopoulos (2000) considered seasonal adjustment based on multiplicative classical decomposition. Moreover, they performed a prior autocorrelation test to determine if the data are seasonal, where the significance level was set to 90%. In this section we will provide some results where we consider

- varying the significance level for identifying a series as seasonal (80, 90, 95%) and
- both additive and multiplicative forms of classical decomposition for seasonal adjustment.

The results are presented in table 1. It is apparent that at least for the M3-Competition monthly series, a multiplicative seasonal adjustment offers by far better results compared to an additive seasonal form. Moreover, we would expect that when dealing with real data, multiplicative seasonal forms occur more naturally. A second observation comes from comparing the various significance levels. Whereas the 80% level is the worse than 90 and 95% for both seasonal forms (multiplicative and additive), the performance of 90 and 95% significance levels depend on the seasonal form. In any case, a 90% significance level coupled with a multiplicative seasonal form provides the best results. Lastly, note that our results slightly deviate from the originally published results (sMAPE of 13.85%). We advocate any differences on the implementation of the method using different software and, consequently, different optimisation techniques.

Further extensions with regards to the prior seasonal adjustment of the data could involve an optimal identification of the seasonal form (additive or multiplicative) so that such a decision is not imposed in aggregate manner (across all time series) but set for each series individually. To the best of our knowledge such a test does not exist, however several heuristics could be applied. Furthermore, one could consider to apply several approaches to robustify the estimation

Table 1: Forecasting performance of the theta method for various seasonal adjustments.

Significance level	Seasonal form	sMAPE (%)
90%	Multiplicative	13.85 [†]
80%	Multiplicative	13.91
90%	Multiplicative	13.87
95%	Multiplicative	13.89
80%	Additive	15.04
90%	Additive	14.96
95%	Additive	14.61

[†]Published result in Makridakis and Hibon (2000).

of the seasonal component. For example, Petropoulos and Nikolopoulos (2013), following the results of (Miller and Williams, 2003), considered possible shrinkage of the seasonal indices. The distribution of the seasonal indices (symmetric or skewed) as well as the value of the James-Stein shrinkage estimator are used to decide between the seasonal indices directly provided by classical decomposition or shrinking these indices by applying the James-Stein or the Lemon-Krutchkoff shrinkage estimators. Petropoulos and Nikolopoulos (2013) report a small improvement in the value of sMAPE (13.78%) if seasonal shrinkage is considered.

4 Optimising the theta lines

Two recent studies, have considered optimisation of the second theta line, the line that focuses on the extrapolation of the short-term behaviour.

Fiorucci et al. (2015) determined the optimal value of this theta line by measuring the performance of the theta method with different lines using a cross-validation exercise. In more detail, they proposed a Generalised Rolling Origin Evaluation (GROE), defined as:

$$l(\theta) = \sum_{i=1}^p \sum_{h=1}^{\min(H, n-t)} g(y_{t+h}, f_{t+h|t}) \quad (2)$$

where y_{t+h} is the actual at period $t+h$, $f_{t+h|t}$ is the h -step-ahead forecast as calculated at period t , H is the number of forecasts produced at each origin, and p denotes how many times the origin is updated. If we assume that m is the step movement for each origin update, with $m \in \mathbb{N}^*$, then

$$p = 1 + \left\lceil \left\| \frac{n - n_1}{m} \right\| \right\rceil, \quad (3)$$

where n_1 is the first origin and $\|x\|$ is a function returning the largest integer so that $\|x\| < x$.

Fiorucci et al. (2015) note that their proposed GROE can be used to construct evaluation schemes with overlapping, semi-overlapping or non-overlapping validation blocks. Also, fixed origin evaluation and rolling origin evaluation as discussed by Tashman (2000), are special cases of the GROE, as follows:

- GROE falls to fixed origin evaluation when $m = H = n - n_1$, which also suggests that $p = 1$.
- GROE falls to rolling origin evaluation when $m = 1$ and $H \geq n - n_1$. Also, the number of origins evaluated can be adjusted by appropriately setting the value of n_1 .

Fiorucci et al. (2015) examined several set ups of their GROE to select between theta models with different θ values so that $\theta \in 1, 1.5, \dots, 4.5, 5$. The θ value of the other theta line was fixed to zero. Fiorucci et al. (2015) referred to this approach as the optimised theta method (OTM), however in this chapter we will refer to it as OTM-GROE.

In an extension of this study, Fiorucci et al. (2016) proposed the following state space (SS) model for the theta method:

$$y_t = \mu_t + \varepsilon_t, \quad (4)$$

$$\mu_t = \ell_{t-1} + \left(1 - \frac{1}{\theta}\right) \left\{ (1 - \alpha)^{t-1} A_n + \left[\frac{1 - (1 - \alpha)^t}{\alpha} \right] B_n \right\}, \quad (5)$$

$$\ell_t = \alpha y_t + (1 - \alpha) \ell_{t-1}, \quad (6)$$

where A_n and B_n are the intercept and slope of the linear regression line as calculated for all observations in the in-sample data and ℓ_0 , α and θ are the parameters of the model, which are estimated by minimising the sum of squared errors. For ℓ_0 and α the usual constraints of SES apply ($\ell_0 \in \mathbb{R}$ and $\alpha \in (0, 1)$) whereas the constrain for the third parameter is $\theta \geq 1$. Fiorucci et al. (2016) referred to the above model as the optimised theta model (OTM); we will hereafter refer to this model as OTM-SS. Note that if $\theta = 2$, then OTM is equivalent to the standard theta method. However, a state space formulation allows for computation of the conditional variance and the corresponding prediction intervals.

Fiorucci et al. (2016) also proposed a dynamic version of the above state space model, so that the coefficients for the intercept and the slope are updated dynamically, along with the other states

of the model. The dynamic optimised theta model (DOTM) is expressed as follows:

$$y_t = \mu_t + \varepsilon_t \quad (7)$$

$$\mu_t = \ell_{t-1} + \left(1 - \frac{1}{\theta}\right) \left[(1 - \alpha)^{t-1} A_{t-1} + \left(\frac{1 - (1 - \alpha)^t}{\alpha}\right) B_{t-1} \right] \quad (8)$$

$$\ell_t = \alpha y_t + (1 - \alpha) \ell_{t-1} \quad (9)$$

$$A_t = \bar{y}_t - \frac{t+1}{2} B_t \quad (10)$$

$$B_t = \frac{1}{t+1} \left[(t-2) B_{t-1} + \frac{6}{t} (y_t - \bar{y}_{t-1}) \right] \quad (11)$$

$$\bar{y}_t = \frac{1}{t} [(t-1) \bar{y}_{t-1} + y_t]. \quad (12)$$

All methods and models in this section, OTM-GROE, OTM-SS and DOTM, implicitly or explicitly consider a linear combination of the actuals and the forecasts of the two theta lines so that the original signal is reconstructed, or:

$$y_t = \omega L_t(\theta_1) + (1 - \omega) L_t(\theta_2). \quad (13)$$

Given that $\theta_1 = 0$ and $\theta_2 = \theta$, Fiorucci et al. (2015) and Fiorucci et al. (2016) show that $\omega = 1 - \frac{1}{\theta}$, which corresponds to the factor multiplied by the linear trend estimation in equations 5 and 8.

Table 2 presents the performance of theta methods and models that consider an optimal θ value for the theta line for the short-term behaviour. More specifically, with regards to the OTM-GROE, two different versions are reported that correspond to the fixed origin and the rolling origin evaluation respectively.

The results in table 2 suggest that optimising the theta line that corresponds to the short-term behaviour (amplified local curvatures) can lead to improved forecasting performance. GROE gives slightly better performance; however, the state space model provide the basis for calculating prediction intervals as well. Also, rolling origin evaluation (cross-validation) does a better job in identifying a optimal θ value compared to fixed-origin evaluation (validation). This was expected, as the former does not over-focus on a single validation window. Moreover, Fiorucci et al. (2016) show that the percentage improvements of DOTM over a non-optimised theta method are greater for trended series compared to non-trended ones.

Table 2: Forecasting performance of the theta method when the θ value of the short-term-behaviour theta line is optimised.

Method/Model	Settings	sMAPE
Standard theta method		13.85 [†]
OTM-GROE	$p = 1, m = H, n_1 = n - H$	13.78 [‡]
OTM-GROE	$p = H, m = 1, n_1 = n - H$	13.66 [‡]
OTM-SS		14.11 [*]
DOTM		13.74 [*]

[†]Published result in Makridakis and Hibon (2000).

[‡]Published result in Fiorucci et al. (2015), sAPE as cost function.

^{*}Published result in Fiorucci et al. (2016).

5 Adding a third theta line

Other research has focused on the performance of the theta method if a third line is added into the mix. For instance, Petropoulos and Nikolopoulos (2013) considered the addition of a third theta line with integer values for the theta parameter of the third theta line so that $\theta \in [-1, 3]$.

- Theta line with $\theta = 1$ corresponds to the deseasonalised series.
- Theta line with $\theta = 0$ is the linear regression on time (no curvatures).
- Theta line with $\theta = 2$ has double the curvatures of the deseasonalised series.
- Theta line with $\theta = 3$ has triple the curvatures of the deseasonalised series.
- Theta line with $\theta = -1$ exhibits curvatures that mirror the curvatures of the deseasonalised series ($\theta = 1$), with theta line with $\theta = 0$ acting as the symmetry axis.

In Petropoulos and Nikolopoulos (2013), the forecast for the third theta line is produced using SES, as is the case for theta line with $\theta = 2$. The optimal value of the third theta line is estimated for each series individually, through maximising the forecasting accuracy on a single validation sample. Also, Petropoulos and Nikolopoulos (2013) varied the contribution weight of this third theta line. The various combinations, together with the corresponding sMAPE values, are reported in table 3 in rows 2 to 6.

We observe that in most of the cases the addition of a third theta line leads to decreases in the value of sMAPE. The best performance is recorded for the case where the weights for lines with $\theta = 0, 2$ and θ are 50, 40 and 10% respectively. The accuracy improvement is, on average, 1.2% compared to the standard theta method.

Table 3: Forecasting performance of the theta method when a third theta line is added.

Function for final forecast	sMAPE (%)
$0.5 \times L(0) + 0.5 \times L(2)$	13.85 [†]
$1/3 \times L(0) + 1/3 \times L(2) + 1/3 \times L(\theta)$	14.34 [‡]
$0.45 \times L(0) + 0.45 \times L(2) + 0.1 \times L(\theta)$	13.71 [‡]
$0.475 \times L(0) + 0.475 \times L(2) + 0.05 \times L(\theta)$	13.70 [‡]
$0.5 \times L(0) + 0.3 \times L(2) + 0.2 \times L(\theta)$	13.74 [‡]
$0.5 \times L(0) + 0.4 \times L(2) + 0.1 \times L(\theta)$	13.68 [‡]
$1/3 \times L(0) + 1/3 \times L(1) + 1/3 \times L(2)$	13.68

[†]Published result in Makridakis and Hibon (2000).

[‡]Published result in Petropoulos and Nikolopoulos (2013).

The last row of table 3 reports a new sets of results, based on the equal-weight combination of the forecasts of three theta lines with $\theta = 0, 1$ and 2 . In this last case, the extrapolation of the additional theta line ($\theta = 1$) is performed using the Damped Exponential Smoothing (DES) method. DES has been proved to be a good benchmark method in several empirical forecasting studies when applied on the seasonally adjusted data. The performance of this simple last combination (which can also be regarded as $2/3 \times$ theta and $1/3 \times$ DES) achieves accuracy similar to the best combination of three theta lines where the θ value of the third line is optimally selected per series ($0.5 \times L(0) + 0.4 \times L(2) + 0.1 \times L(\theta)$).

In any case, more than three theta lines might be also produced and calculated separately. Theta lines with θ values lower than unity are better for estimating the long-term component, while theta lines with θ values higher than unity focus on the short term curvatures.

6 Adding a short-term linear trend line

The standard theta line considers a single linear trend that is calculated using all data points of the seasonally adjusted series. However, as it is often in practice, the trend of a time series changes over time, so that the long-term trend and the short-term trend may exhibit different direction. In such cases, a theta method that is simply based on the long-term trend might be an over- or under-estimation of the short-term reality.

We suggest that theta method could be expanded so that a second linear trend line is added. This second linear trend line should focus on the short-term trend component. In other words, a linear trend is fitted on the last k data points of the seasonally adjusted data. There are several

options for specifying the value of k , however two simple alternatives can be linked with the problem and data in hand:

- $k = H$, where H is the required forecasting horizon and $H < n$.
- $k = m$, where m is the periodicity of the data ($m = 12$ for monthly data) and $m < n$.

Table 4 presents the performance of theta method when a short-term linear trend line is added. It appears that for the specific case of the monthly M3-competition data, the sMAPE is a U-shaped function of k and that a minimum exists for $k = H$, where the value of sMAPE is 13.68%. We conclude that the addition of a linear line that focuses on the short-term trend can add value on the theta method.

Table 4: Forecasting performance of the theta method when a short-term linear trend line is added.

Short-term trend estimation window (k)	sMAPE (%)
—	13.85 [†]
m	14.49
H	13.68
$2 \times m$	13.99
$2 \times H$	14.40

[†]Published result in Makridakis and Hibon (2000).

7 Extrapolating theta lines

Originally, Assimakopoulos and Nikolopoulos (2000) suggested that the theta line with $\theta = 0$, corresponding to the linear regression line, should be extrapolated as usual assuming linear trend for the future values. At the same time, the theta line with $\theta = 2$ should be forecasted using SES. However, alternative extrapolating methods for these two lines may be considered.

Hyndman et al. (2002) proposed a state space framework for the exponential smoothing family of models. Each model consists of three terms, the error term (which can be either additive, A, or multiplicative, M), the trend component (which can be either none, N, additive, A, or multiplicative, M) and the seasonal term (which similar to trend can be either none, N, additive, A, or multiplicative, M). Additionally, if trend exists, this might be damped or not. Under this framework, an exponential smoothing (ETS) model form is abbreviated with these three terms. For example, ETS(M,Ad,N) is an exponential smoothing model with multiplicative errors, additive damped trend and no seasonality. Along with this framework, Hyndman et al. (2002) proposed

how automatic forecasting should be applied with the use of information criteria, which account for the goodness-of-fit of the candidate models after penalising it to take into account the varying complexity across different models (number of parameters). The model with the minimum value for a pre-specified information criterion is selected as the optimal model.

Another alternative for automatic forecasting would be to use the Autoregressive Integrated Moving Average (ARIMA) family of models. This involves the correct identification of the autoregressive order (AR, $p \in \mathbb{N}$), differencing order ($d \in \mathbb{N}$) and moving average order (MA, $q \in \mathbb{N}$). Obviously, infinite number of combinations and potential models exist, however high-order models lead to overfitting. Hyndman and Khandakar (2008) proposed an algorithm that, after applying the appropriate number of differences as suggested by unit-root tests, initially considers a set of simple ARIMA models (ARIMA(2,d,2), ARIMA(0,d,0), ARIMA(1,d,0) and ARIMA(0,d,1)) and selects the one with the lowest value on the corrected Akaike's information criterion (AICc). Subsequently, a stepwise search is performed so that the AR and MA orders are modified by one unit (plus or minus) until a better model cannot be found. Moreover, the default implementation of the above algorithm (hereafter referred to as AutoARIMA), which is available as the `auto.arima()` function of the `forecast` package of the R statistical software, limits the maximum number of AR and MA terms to five so that overfitting is further avoided.

Either we consider ETS or AutoARIMA as automatic forecasting frameworks, it is apparent that these can be applied to produce forecasts for any of the theta lines with $\theta \neq 0$. We assume the three-line theta method that consists of the theta lines with $\theta = 0, 1$ and 2 (section 5) and we apply ETS and ARIMA to either or both theta lines with $\theta = 1$ and $\theta = 2$. We benchmark against the standard two-line theta method as well as the three-line theta method where the theta lines with $\theta = 1$ and $\theta = 2$ have been extrapolated by Damped and SES respectively.

Table 5 presents the respective results. Interestingly, applying automatic forecasting approaches to extrapolate either of the theta lines does not lead to improvements. Simply extrapolating the theta lines with fixed and pre-specified methods performs best. In any case, the performance of all combinations tested is very similar, apart from the case where AutoARIMA is applied to both lines with $\theta = 1$ and $\theta = 2$.

Table 5: Forecasting performance of the theta method using automatic forecasting for extrapolating theta lines with $\theta \neq 0$.

$\theta = 1$	$\theta = 2$	sMAPE (%)
—	SES	13.85 [†]
Damped	SES	13.68
ETS	SES	13.68
Damped	ETS	13.78
ETS	ETS	13.85
AutoARIMA	SES	13.87
Damped	AutoARIMA	13.82
AutoARIMA	AutoARIMA	14.24
ETS	AutoARIMA	13.90
AutoARIMA	ETS	13.96

[†]Published result in Makridakis and Hibon (2000).

8 Combination weights

The standard theta method suggests that once the two theta lines (with $\theta = 0$ and $\theta = 2$) are extrapolated, the two sets of forecasts should be combined with equal weights. However, Petropoulos and Nikolopoulos (2013) considered optimising these weights for each series separately, based on the out-of-sample performance on a hold-out-sample consisting of the last year of data (12 observations). In other words, they optimised, through validation, the value of ω when $y_t = \omega L_t(\theta_1) + (1 - \omega)L_t(\theta_2)$. Moreover, they examined several parametric spaces for the value of ω , so that ω deviates from 10% up to 40% from equal-weight combination (corresponding to $\omega = 0.5$).

Table 6 presents the respective results. Small deviations from the 50-50% combination enhance the performance of the theta method. For $\omega \in [0.45, 0.55]$, the average value of sMAPE drops to 13.65%. Smaller improvements are recorded as the deviations from the equal weights increase, while larger deviations lead to small decrease in the performance compared to an equal-weight combination. It should be noted, however, that considering non-equal weights for combining the forecasts for theta lines with symmetric values (as is the case for theta lines with $\theta = 0$ and $\theta = 2$) results in a signal that is not a “reconstruction” of the seasonally-adjusted series (theta line with $\theta = 1$).

An alternative approach to combining with unequal weights the forecasts produced through extrapolating the two theta lines derives from the nature of the theta lines itself. In the original study, Assimakopoulos and Nikolopoulos (2000) suggest that when $\theta < 1$ then the long-term

Table 6: Forecasting performance of the theta method with optimised combination weights.

Weights (range)	sMAPE (%)
Equal weights	13.85 [†]
[0.45, 0.55]	13.65 [‡]
[0.4, 0.6]	13.70 [‡]
[0.35, 0.65]	13.83 [‡]
[0.3, 0.7]	14.00 [‡]

[†]Published result in Makridakis and Hibon (2000).

[‡]Published result in Petropoulos and Nikolopoulos (2013).

behaviour of the series is amplified. At the same time, $\theta > 1$ results in theta lines with enhanced short-term behaviour of the data. Following this logic, one could consider weights that vary as the combined forecast for different horizons is computed. In more detail, a higher weight should be assigned on the theta line with $\theta = 2$ for the shorter horizons, while a lower weight should be assigned for the longer horizons.

Assuming that ω deviates within a particular range $[0.5 - d, 0.5 + d]$, we consider a linear function of ω so that $\omega(h) = (0.5 - d) + \frac{2d(h-1)}{(H-1)}$, where h is the forecast horizon and H is the maximum forecast horizon ($H = 18$ for the data used in this chapter). Table 7 presents the results for this approach. As it is evident, simply setting the weights of the theta lines forecasts to match the short- or long-term horizons does not improve the average performance of the theta method. On the contrary, as the value of d increases, the value of sMAPE increases as well.

Table 7: Forecasting performance of the theta method with combination weights suggested by the forecasting horizon.

Weights (range)	sMAPE (%)
Equal weights	13.85 [†]
[0.45, 0.55]	13.94
[0.4, 0.6]	14.07
[0.35, 0.65]	14.25
[0.3, 0.7]	14.47

[†]Published result in Makridakis and Hibon (2000).

9 A robust theta method

Building on the results from the previous sections, we propose a simple and robust modification of the standard theta method that adds

- a third theta line, which is extrapolated with Damped exponential smoothing (section 5) and

- a short-term linear trend line (section 6).

Table 8 presents the results of the robust theta method as applied on the monthly subset of the M3-competition data. The robust theta method reduces the average value of sMAPE to 13.59%, which is the lowest compared to all modifications suggested in this chapter.

Table 8: Forecasting performance of the standard and robust theta methods.

Method	sMAPE (%)
Standard theta method	13.85 [†]
Robust theta method	13.59

[†]Published result in Makridakis and Hibon (2000).

10 Applying theta method in R statistical software

There are several packages in R that provide different implementations of the theta method (and some of its variations).

- `stheta` function of the `forecTheta` package (version 2.2) is to the best of our knowledge the closest open source implementation to the standard theta method as it was applied to the M3-competition.
- `stm` and `dstm` functions of the `forecTheta` package (version 2.2) are implemented based on the state space model proposed by Fiorucci et al. (2016). The main difference between `stm` and `dstm` is that the latter dynamically updates the values of the intercept and slope coefficients of the theta line with $\theta = 0$. The user can define the initialisation of the parameters, the parameter search space as well as the optimisation method to be used (selection between Nelder-Mead, L-BFGS-B or SANN). The function can output prediction intervals as well, where the confidence levels can be specified by the user.
- `otm` and `dotm` functions of the `forecTheta` package (version 2.2) work similar to the functions `stm` and `dstm` with one important difference. They introduce an additional parameter, which is the θ value for the theta line focusing on the short-term curvatures ($\theta > 1$). In other words, the `stm` and `dstm` are special cases of `otm` and `dotm`, if the value of θ is fixed to 2. The value for the other theta line is kept fixed, $\theta = 0$. Moreover, the combination weights for the two theta lines are calculated so that the original signal is reproduced. The respective models are described in Fiorucci et al. (2016).

- `otm.arxiv` function of the `forecTheta` package (version 2.2) allows for optimal selection of the θ value for the second theta line (similarly to the `otm` function). The main difference between `otm` and `otm.arxiv` is that while the former performs optimisation via maximising the likelihood of the respective state space model, the latter does not consider θ as an additional parameter within the model. Instead, it selects the optimal θ value through a cross-validation exercise. Regarding this, the user can specify the cost function, the first origin, the number of origins skipped in each step, the number of origin updates and the forecast horizon. The respective algorithm and parameters are discussed in detail by Fiorucci et al. (2015). However, this function cannot output prediction intervals of the produced forecasts.
- `thetaf` function of the `forecast` package (version 8.2) provides forecasts from the SES with drift method (Hyndman and Billah, 2003). Data are tested for seasonality and adjusted along the lines of Assimakopoulos and Nikolopoulos (2000). The function can also provide prediction intervals of desired levels, which are computed based on the state space model of Hyndman and Billah (2003).
- `theta` function of the `TStools` package is an implementation of the standard theta method with some tweaks. First, the existence of trend and seasonality is done automatically, where the significance level can be specified by the user. Second, it allows for both additive and multiplicative seasonality. Third, the function supports alternative cost functions to the mean squared error for the estimation of the theta lines and the seasonal element. Lastly, it allows for exclusion of outliers for the estimation of theta line with $\theta = 0$, however their timings have to be specified by the user.

Petropoulos and Nikolopoulos (2017) provide a step-by-step tutorial of how the standard version of theta method can be applied in practice. They also show how this can be done in R statistical software using just 10 lines of code.

References

- Assimakopoulos, V., Nikolopoulos, K., 2000. The Theta model: a decomposition approach to forecasting. *International Journal of Forecasting* 16 (4), 521–530.
- Fiorucci, J. A., Pellegrini, T. R., Louzada, F., Petropoulos, F., 2015. The optimised theta method. Arxiv Working Paper <http://arxiv.org/abs/1503.03529>.

- Fiorucci, J. A., Pellegrini, T. R., Louzada, F., Petropoulos, F., Koehler, A. B., 2016. Models for optimising the theta method and their relationship to state space models. *International Journal of Forecasting* 32 (4), 1151–1161.
- Goodwin, P., Lawton, R., 1999. On the asymmetry of the symmetric MAPE. *International Journal of Forecasting* 15 (4), 405–408.
- Hyndman, R. J., Billah, B., 2003. Unmasking the theta method. *International Journal of Forecasting* 19, 287–290.
- Hyndman, R. J., Khandakar, Y., 2008. Automatic time series forecasting: The forecast package for R. *Journal of Statistical Software* 27 (3), 1–22.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., Grose, S., 2002. A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of Forecasting* 18 (3), 439–454.
- Makridakis, S., Hibon, M., 2000. The M3-competition: results, conclusions and implications. *International Journal of Forecasting* 16 (4), 451–476.
- Miller, D. M., Williams, D., Oct. 2003. Shrinkage estimators of time series seasonal factors and their effect on forecasting accuracy. *International Journal of Forecasting* 19 (4), 669–684.
- Petropoulos, F., Nikolopoulos, K., 2013. Optimizing theta model for monthly data. *Proceedings of 5th International Conference on Agents and Artificial Intelligence ICAART 2013*.
- Petropoulos, F., Nikolopoulos, K., 2017. The theta method. *Foresight* 45 (Summer 2017), 11–17.
- Tashman, L. J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* 16 (4), 437–450.